

METODE DEEP LEARNING DALAM TEKNOLOGI DEEPPFAKE : SYSTEMATIC LITERATURE REVIEW

Yoggy Arif Fernandes, Yulia Fatma

Teknik Informatika, Universitas Muhammadiyah Riau
190401062@student.umri.ac.id

ABSTRAK

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) telah membawa dampak signifikan dalam berbagai bidang, terutama dalam pemrosesan citra dan video. Salah satu fenomena yang muncul akibat kemajuan ini adalah *Deepfake*, yaitu teknologi berbasis AI yang memungkinkan manipulasi wajah dan suara secara realistis. Meskipun memiliki potensi positif dalam industri kreatif dan pendidikan, *Deepfake* juga menimbulkan permasalahan serius, seperti penyebaran informasi palsu, ancaman terhadap privasi, dan penyalahgunaan untuk tujuan kriminal. Penelitian ini bertujuan untuk meninjau secara sistematis metode deteksi *Deepfake* menggunakan pendekatan *Deep Learning*. Kajian ini menganalisis 15 artikel ilmiah dari jurnal terakreditasi dan internasional yang diterbitkan antara tahun 2019 hingga 2024. Metode yang dikaji meliputi penggunaan *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM), dan *Generative Adversarial Networks* (GAN) dalam mendeteksi konten manipulatif. Metode penelitian yang digunakan dalam tinjauan literatur ini mencakup pencarian dan seleksi artikel berdasarkan kriteria tertentu, analisis terhadap teknik deteksi *Deepfake*, serta evaluasi dataset yang digunakan dalam penelitian sebelumnya. Hasil tinjauan menunjukkan bahwa CNN dan LSTM banyak diterapkan dalam analisis pola visual dan temporal pada video *Deepfake*, sementara GAN sering digunakan untuk mengembangkan model deteksi yang lebih adaptif. Meskipun berbagai metode deteksi telah dikembangkan, tantangan masih tetap ada, terutama dalam menghadapi peningkatan kualitas *Deepfake* yang semakin sulit dibedakan dari video asli. Oleh karena itu, penelitian lebih lanjut diperlukan untuk meningkatkan keandalan sistem deteksi *Deepfake*. Kajian ini diharapkan dapat memberikan wawasan bagi peneliti dan praktisi dalam mengembangkan solusi yang lebih efektif dalam menangani ancaman *Deepfake* di masa depan.

Kata kunci : *Deep Learning, DeepFake, Computer Vision, Artificial Intelligence*

1. PENDAHULUAN

Saat ini, perkembangan teknologi sangat pesat, terutama dalam bidang Kecerdasan Buatan (AI). AI banyak digunakan dalam teknologi pintar untuk menciptakan sistem otomatis. Selain itu, AI juga efektif dalam pengolahan citra dan video, memberikan hasil yang memuaskan. Namun, penting untuk diingat bahwa tidak semua teknologi yang menggunakan AI memiliki dampak positif; ada juga potensi dampak negatif yang perlu diperhatikan.

Salah satu dampak negatif yang muncul dari penggunaan kecerdasan buatan (AI) adalah fenomena *Deepfake*, yang memungkinkan pembuatan video atau gambar yang sangat realistis menggunakan metode *Deep Learning* (DL) [1].

Dengan teknologi *Deepfake*, seseorang bisa melakukan pertukaran wajah dan mengubah berbagai atribut fisik seperti gaya rambut, jenis kelamin, dan usia, sehingga menghasilkan konten visual yang sulit dibedakan dari yang asli.

Awalnya, teknologi ini digunakan dalam aplikasi seperti *FaceApp* dan *FakeApp* untuk melakukan pertukaran wajah pada gambar dan video dengan hasil yang tampak natural [2].

Namun, seiring dengan perkembangannya, muncul kekhawatiran mengenai penggunaannya untuk tujuan yang tidak etis, seperti menyebarkan informasi

palsu atau membuat video pornografi, yang dapat mengancam privasi dan keamanan digital. Penyebaran video *Deepfake* yang melibatkan wajah selebritas menimbulkan reaksi negatif dari masyarakat, dan banyak platform media sosial serta situs dewasa mulai melarang konten semacam ini [3].

Menanggapi ancaman yang ditimbulkan oleh *Deepfake*, banyak upaya deteksi yang dikembangkan oleh akademisi dan pakar teknologi. Metode deteksi yang menggunakan *Deep Learning* telah mendapatkan perhatian luas dalam beberapa tahun terakhir, dengan banyak penelitian yang fokus pada pengembangan teknik yang lebih efektif dalam mengidentifikasi video atau gambar palsu [4].

Meskipun demikian, para peneliti masih menghadapi tantangan besar karena teknologi *Deepfake* terus berkembang dan semakin canggih.

Dengan latar belakang tersebut, tinjauan literatur ini bertujuan untuk memberikan pemahaman menyeluruh tentang perkembangan metode *Deep Learning* dalam konteks *Deepfake*, terutama terkait teknik deteksinya. Dalam tinjauan ini, akan membahas berbagai metode yang telah diusulkan, *dataset* yang digunakan dalam pelatihan model deteksi. Kami berharap tinjauan ini dapat memberikan wawasan yang berguna bagi peneliti dan praktisi yang tertarik dengan teknologi *Deepfake* dan deteksi konten palsu.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian terdahulu membahas pendekatan deep learning dalam mendeteksi deepfake. meneliti efektivitas CNN dan Transformer dalam mendeteksi konten manipulatif, dengan hasil menunjukkan bahwa meskipun CNN unggul dalam ekstraksi fitur visual, Transformer mampu memberikan akurasi lebih tinggi dalam beberapa kasus [5].

Sementara itu, penelitian lain-nya mengevaluasi empat model pre-trained—XceptionNet, InceptionResNetV2, EfficientNetV2S, dan EfficientNetV2M—pada dataset FaceForensics++ dan CelebDf-V2, dengan temuan bahwa EfficientNetV2M memiliki kinerja terbaik dalam mendeteksi wajah deepfake secara presisi[6].

Penelitian[7] mengusulkan kombinasi CNN dan Vision Transformer untuk meningkatkan akurasi deteksi video deepfake. Model ini diuji pada dataset DFDC dan mencapai akurasi 91,5%, menunjukkan efektivitas pendekatan gabungan dalam mengidentifikasi video yang telah dimanipulasi.

2.2. Computer Vision

Computer Vision (CV) merupakan cabang dari ilmu komputer yang berfokus pada cara komputer dapat mengenali dan menginterpretasikan objek, individu, serta konten visual lainnya dalam bentuk gambar atau video. CV adalah salah satu aplikasi kecerdasan buatan (AI) yang memanfaatkan teknologi AI dan *Machine Learning* (ML) untuk mengolah data visual. Teknologi ini meniru kemampuan manusia dalam mengamati dan memahami apa yang dilihat. Beberapa tugas utama yang dapat dilakukan oleh CV meliputi pengenalan objek dan wajah, pembacaan teks, serta pemahaman konteks dari gambar atau video[8].

2.3. Deep Learning

Deep learning adalah salah satu cabang dari *machine learning* yang memungkinkan komputer untuk belajar dari pengalaman dan memahami dunia melalui hierarki konsep. Karena komputer memperoleh pengetahuan dari pengalaman, manusia tidak perlu secara rinci menentukan apa yang perlu diketahui oleh komputer. Konsep hierarki ini memungkinkan komputer untuk mempelajari hal-hal yang kompleks dengan cara menyederhanakannya; ini menciptakan sebuah graf yang terdiri dari banyak lapisan yang dalam atau "*deep*." [9]

2.4. Deepfake

Deepfake merupakan gabungan dari kata "*deep learning*" dan "*fake*." Ini adalah video atau citra yang dimanipulasi secara digital sehingga terlihat sangat realistis, menampilkan seseorang seolah-olah sedang berbicara atau melakukan sesuatu yang sebenarnya tidak pernah terjadi. *Deepfake* menggunakan *neural network* yang menganalisis kumpulan data besar untuk mempelajari cara meniru ekspresi wajah, perilaku, dan

bahkan suara seseorang. Proses ini melibatkan penginputan dua gambar dari orang yang sama ke dalam algoritma *deep learning* untuk melatihnya dalam menukar wajah. Dengan kata lain, *deepfake* memanfaatkan teknologi pemetaan wajah dan AI untuk mengganti wajah satu orang dengan wajah orang lain dalam sebuah video atau citra [10]

2.5. Forensic Deep Learning

Forensics Deep Learning adalah metode yang digunakan untuk menentukan apakah sebuah gambar asli atau telah dimanipulasi. Proses ini dilakukan dengan teknik forensics yang bertujuan untuk mengukur seberapa besar perubahan pada gambar menggunakan *error level analysis* (ELA). ELA sendiri adalah teknik yang menghitung perbedaan nilai *luminance* (kecerahan) dan *chrominance* (warna) untuk melihat seberapa banyak gambar tersebut telah diubah. Hasil dari analisis ini biasanya menghasilkan gambar dengan area yang dominan hitam dan titik-titik putih, yang menunjukkan bagian asli atau yang telah dimanipulasi[11].



Gambar 1. Contoh Deepfake

3. METODE PENELITIAN

Penelitian ini menggunakan metode Systematic Literature Review (SLR) sebagai pendekatan utama dalam mengumpulkan dan menganalisis data. Proses penelitian dilakukan secara sistematis melalui beberapa tahapan, yaitu perencanaan, pencarian literatur, seleksi sumber, ekstraksi data, dan analisis sintesis.

3.1. Perencanaan dan Identifikasi Topik

Tahap awal penelitian dimulai dengan perumusan pertanyaan penelitian yang berfokus pada metode deteksi deepfake berbasis deep learning. Tujuan utama dalam tahap ini adalah mengidentifikasi tren perkembangan teknologi deepfake serta metode deteksi yang telah diterapkan pada berbagai studi sebelumnya.

3.2. Pencarian dan Seleksi Literatur

Sumber literatur diperoleh dari berbagai database akademik yang kredibel, termasuk Google Scholar, IEEE Xplore, Elsevier, SpringerLink, Cornell University Library, dan ACM Digital Library. Kriteria inklusi dalam pemilihan literatur mencakup:

- Artikel jurnal atau konferensi yang telah dipublikasikan dalam rentang waktu 2019 hingga 2024.
- Studi yang membahas metode deteksi deepfake dengan pendekatan deep learning seperti CNN, GAN, LSTM, Transformer, dan metode hibrida lainnya. Penelitian yang menggunakan dataset populer seperti FaceForensics++, Celeb-DF, DeepFake Detection Challenge (DFDC), atau dataset lainnya yang sering digunakan dalam evaluasi model.

Sementara itu, kriteria eksklusi mencakup artikel yang:

- Tidak membahas metode berbasis deep learning. Tidak memiliki informasi yang cukup mengenai metode, dataset, atau hasil eksperimen.
- Merupakan preprint yang belum melalui proses peer review.

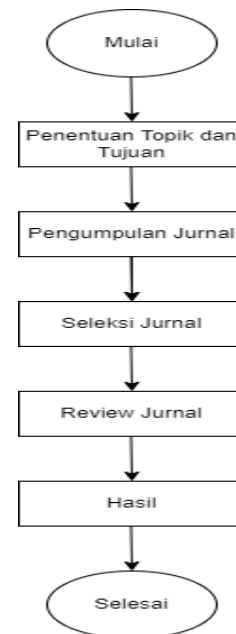
Dari hasil pencarian awal, ditemukan lebih dari 50 artikel yang sesuai dengan kata kunci yang digunakan. Namun, setelah proses seleksi lebih lanjut berdasarkan abstrak, metode, dan hasil penelitian, hanya 15 artikel yang dianggap relevan dan masuk dalam kajian utama.

3.3. Sintesis Temuan dan Penyusunan Kesimpulan

Tahap akhir adalah melakukan sintesis terhadap hasil analisis untuk memperoleh pemahaman yang lebih mendalam. Hasil studi ini tidak hanya memberikan ringkasan metode deteksi deepfake yang telah digunakan dalam penelitian sebelumnya, tetapi juga memberikan wawasan mengenai potensi pengembangan model di masa depan.

Gambar 2 Penelitian ini dimulai dengan menentukan topik dan tujuan, yaitu mengevaluasi metode deep learning dalam deteksi deepfake. Setelah itu, jurnal dikumpulkan dari berbagai sumber akademik terpercaya menggunakan kata kunci yang

relevan. Selanjutnya, jurnal yang diperoleh diseleksi berdasarkan kriteria tertentu, seperti rentang tahun publikasi, metode yang digunakan, serta relevansi dengan topik penelitian. Setelah jurnal yang sesuai terpilih, dilakukan analisis mendalam terhadap metode, dataset, serta parameter evaluasi yang digunakan dalam penelitian sebelumnya. Hasil analisis ini kemudian dibandingkan untuk mengidentifikasi tren terbaru serta tantangan yang masih dihadapi dalam deteksi deepfake. Tahap akhir penelitian ini adalah menyusun kesimpulan berdasarkan temuan yang diperoleh. Dengan sintesis dari berbagai jurnal, penelitian ini diharapkan memberikan wawasan berharga bagi akademisi dan praktisi yang tertarik pada teknologi deteksi deepfake



Gambar 2. Metode Penelitian

4. HASIL DAN PEMBAHASAN

Tabel 1. perbandingan literatur review

No	Judul Artikel	Metode	Jumlah Dataset	Akurasi	Objek
1	A GAN based Model of Deepfake Detection in Social Media (2023)	Generative Adversarial Networks	100000	100%	Gambar
2	Trans-DF: A Transfer Learning-based end-to-end Deepfake Detector (2020)	Generative Adversarial Networks (GANs)	9000	90.2%	Video
3	Spatio-temporal Features for Generalized Detection of Deepfake Videos (2020)	CNN 3D	1000	78.57%	Gambar
4	WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection (2024)	CNN dan RNN	7314	85.50%	Gambar
5	Deepfake Video Detection through Optical Flow based CNN (2019)	Optical Flow dan CNN	1000	VGG16: 81.61% ResNet50: 75.46%	Video
6	Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network (2024)	Convolutional Neural Network (CNN)	80000	87.06%	Gambar
7	Deepfake detection in videos using Long Short-Term Memory and CNN ResNext (2022)	ResNext CNN dan LSTM	10000	90% (40, 60 frame)	Video

No	Judul Artikel	Metode	Jumlah Dataset	Akurasi	Objek
8	A Convolutional LSTM based Residual Network for Deepfake Video Detection (2020)	CLRNet dengan Convolutional LSTM	8363	83.08% (FaceSwap)	Video
9	A Note on Deepfake Detection with Low-Resources (2020)	MesoNet	19000	86.89% (MesoNet)	Gambar
10	Fighting Deepfake by Exposing the Convolutional Traces on Images (2020)	Expectation-Maximization	2000	89% - 98%	Gambar
11	Sharp Multiple Instance Learning for (2020)	Sharp Multiple Instance Learning (S-MIL)	1000	83.78% (FFPMS)	Video
12	DeepFake Detection Based on Discrepancies Between Faces and their Context (2020)	XceptionNet	1000	AUC 82,57%	Gambar
13	Analisis Gambar Wajah Palsu: Mendeteksi Keaslian Gambar Yang Dimanipulasi Menggunakan Metode Variational Autoencoder Dan Forensics Deep Neural Network (2023)	Variational Autoencoder (VAE)	202.599	93,17%	Gambar
14	Dynamic texture analysis for detecting fake faces in video sequences (2021)	Support Vector Machine (SVM)	1000	>90% (DF, FSW)	Video
15	Deepfake-Image Anti-Forensics with Adversarial Examples Attacks (2021)	Poisson noise DeepFool (PNDF)	10000	AUC 99%	Gambar

4.1. A GAN-Based Model of Deepfake Detection in Social Media

Penelitian ini [12] menggunakan Algoritma *Deep Convolutional GAN* (DCGAN). Algoritma ini digunakan untuk mendeteksi gambar deepfake dengan melakukan pembangkitan gambar palsu yang realistis dan mendeteksi apakah gambar tersebut asli atau palsu. Model DCGAN mampu mendeteksi *deepfake* dengan tingkat akurasi hingga 100% pada iterasi ke 10 dari 8000 iterasi (epoch). Tidak ada informasi spesifik mengenai waktu yang diperlukan dalam proses pelatihan.

Optimasi dilakukan dengan menggunakan *Binary Cross-Entropy* (BCE) sebagai fungsi loss dan metode *Adam* sebagai optimizer. *Adam* (Adaptive Moment Estimation) adalah algoritma optimasi yang menggabungkan keuntungan dari dua metode optimasi lainnya: momentum dan RMSprop. *Adam* bekerja dengan menghitung momen pertama (mean) dan momen kedua (variance) dari gradien, sehingga proses pembelajaran menjadi lebih cepat dan stabil, terutama pada data yang berisik atau skenario dengan gradien yang jarang berubah.

Dalam proses pelatihan GAN, data input yang diberikan kepada generator disebut data acak z . Ini merupakan vektor acak yang diambil dari distribusi tertentu, biasanya distribusi Gaussian atau seragam, yang digunakan sebagai titik awal untuk menghasilkan gambar. Vektor acak z biasanya berbentuk vektor 1 dimensi dengan panjang yang telah ditentukan, seperti 100 dimensi, dan berfungsi sebagai masukan untuk jaringan generator. Generator kemudian mengubah vektor acak ini menjadi gambar yang realistis melalui serangkaian operasi konvolusi terbalik (transpose convolution).

Arsitektur DCGAN ini menggunakan lapisan konvolusi tanpa pooling dan melakukan upsampling melalui konvolusi transpose. Tidak ada penambahan arsitektur yang signifikan disebutkan, tetapi

arsitektur yang sudah dirancang untuk optimasi tugas pembangkitan dan deteksi gambar

4.2. Deepfake detection in videos using Long Short-Term Memory and CNN ResNext

Penelitian ini [13] menggunakan kombinasi *ResNext Convolutional Neural Network* (ResNext CNN) dan *Long Short-Term Memory* (LSTM). ResNext CNN digunakan untuk proses ekstraksi fitur, sedangkan LSTM digunakan untuk proses klasifikasi data video *deepfake*. Akurasi tertinggi yang dicapai adalah 90% pada data dengan 40 dan 60 frame. Untuk data dengan 10 frame, akurasi yang dihasilkan lebih rendah yaitu hanya 52%. Tidak disebutkan secara eksplisit durasi waktu pelatihan model dan tidak ada juga informasi khusus tentang jumlah *epoch* yang digunakan dalam proses pelatihan dalam penelitian ini.

Model menggunakan kombinasi CNN ResNext dan LSTM, tetapi tidak disebutkan secara eksplisit apakah ada teknik optimasi khusus yang digunakan selain pendekatan yang umum seperti penggunaan *confusion matrix* untuk evaluasi performa model. Arsitektur yang menggabungkan ResNext CNN untuk ekstraksi fitur dari video dan LSTM untuk klasifikasi. Penggunaan ResNext merupakan optimasi dari CNN yang memberikan kinerja lebih baik dalam pengolahan data. Namun, arsitektur tambahan atau modifikasi lain tidak disebutkan secara eksplisit.

4.3. A Note on Deepfake Detection with Low-Resources

Penelitian ini [14] mengusulkan dua algoritma utama yaitu *MesoNet*, yang dioptimalkan dengan mengganti fungsi aktivasi *ReLU* dan *LeakyReLU* dengan fungsi aktivasi baru bernama *Pish*. *Local Feature Descriptors* (LFD), yang menggunakan metode deskripsi fitur lokal untuk mendeteksi deepfake tanpa memerlukan GPU. *MesoNet* dengan kombinasi fungsi aktivasi *Pish* dan *Mish* mencapai

akurasi sebesar 0,8689. LFD dengan metode berbasis BRISK memberikan hasil akurasi dan *recall* yang melebihi 0,7. LFD memiliki waktu inferensi lebih cepat dibandingkan *MesoNet* ketika digunakan pada perangkat dengan daya komputasi terbatas.

Tidak disebutkan secara eksplisit jumlah *epoch* yang digunakan dalam penelitian ini. *MesoNet* menggunakan optimasi melalui penggantian fungsi aktivasi dari *ReLU* dan *LeakyReLU* dengan *Pish* dan *Mish*, yang meningkatkan akurasi dan konsistensi. LFD menggunakan metode sederhana tanpa GPU, jadi tidak ada teknik optimasi yang disebutkan secara eksplisit.

MesoNet menggunakan arsitektur yang dioptimalkan dengan mengganti fungsi aktivasi standar dengan *Pish* dan *Mish*, yang memberikan peningkatan performa. LFD menggunakan pendekatan yang lebih ringan tanpa memerlukan GPU, sehingga lebih sederhana dalam implementasi dibandingkan *MesoNet*.

4.4. Fighting Deepfake by Exposing the Convolutional Traces on Images

Penelitian ini [15] menggunakan algoritma *Expectation-Maximization* yang dilatih untuk mendeteksi dan mengekstrak sidik jari yang dikenal sebagai Jejak Konvolusi (*Convolutional Traces/CT*) yang ditinggalkan oleh GAN selama proses pembuatan gambar. Akurasi klasifikasi keseluruhan lebih dari 98%, yang mempertimbangkan *Deepfake* dari 10 arsitektur GAN yang berbeda. Pada *Deepfake* yang dihasilkan oleh FACEAPP, akurasi mencapai 93%. Tidak ada informasi eksplisit mengenai durasi waktu yang diperlukan dalam pelatihan atau inferensi model. Informasi mengenai jumlah epoch tidak disebutkan dalam penelitian ini.

Algoritma menggunakan *Expectation-Maximization* sebagai bagian dari metode optimasi untuk mendeteksi dan mengekstrak *Convolutional Traces* (CT) dari gambar yang dihasilkan oleh GAN. Ini merupakan pendekatan yang dioptimalkan untuk deteksi *Deepfake*. Arsitektur metode ini menggabungkan Jejak Konvolusi (CT) yang diekstrak menggunakan *Expectation-Maximization*, yang berbeda dari metode deteksi *Deepfake* pada umumnya.

4.5. Sharp Multiple Instance Learning for DeepFake Video Detection

Penelitian ini [16] menggunakan algoritma *Sharp Multiple Instance Learning* (S-MIL), yang merupakan pendekatan baru untuk deteksi video *deepfake*. S-MIL membuat prediksi langsung dari *embedding instance* ke prediksi keseluruhan video tanpa perlu membuat prediksi instance individu terlebih dahulu. Akurasi dari metode S-MIL pada beberapa dataset adalah: DFDC: 0.8378 Celeb-DF: 0.9923 FFPMS: 0.9036. Metode lain yang dibandingkan dengan S-MIL memiliki akurasi yang lebih rendah pada dataset yang sama, misalnya pada DFDC (0.8107), Celeb-DF (0.8609), dan FFPMS (0.7857).

Tidak ada informasi eksplisit mengenai waktu pelatihan atau inferensi dalam penelitian ini. Informasi mengenai jumlah *epoch* yang digunakan dalam pelatihan tidak disebutkan secara spesifik. Arsitektur S-MIL memiliki tambahan pengkodean spasial-temporal yang dirancang untuk menangkap inkonsistensi antar frame video yang dimanipulasi, meningkatkan akurasi deteksi manipulasi parsial pada video *deepfake*.

4.6. DeepFake Detection Based on Discrepancies Between Faces and their Context

Penelitian ini [17] menggunakan algoritma *Xceptionet*. Pengujian dilakukan menggunakan AUC (Area Under the Curve), dan metode ini mencapai AUC 0.8257, menunjukkan kemampuan yang baik dalam mendeteksi *deepfake*, termasuk pada metode manipulasi yang belum pernah dilihat sistem sebelumnya. Tidak disebutkan secara spesifik informasi mengenai waktu pelatihan atau inferensi untuk sistem ini. Tidak ada informasi khusus terkait jumlah *epoch* yang digunakan dalam pelatihan jaringan ini.

Meskipun tidak disebutkan secara eksplisit, metode ini menggunakan optimasi dalam bentuk deteksi artefak spesifik dari metode manipulasi wajah tertentu, yang membantu meningkatkan keandalan deteksi. Arsitektur ini memiliki tambahan dalam bentuk dua jaringan berbeda—jaringan pengenalan wajah dan jaringan pengenalan konteks—yang bekerja bersama-sama untuk mendeteksi manipulasi dengan menganalisis perbedaan antara area wajah yang dimanipulasi dan area yang tidak dimanipulasi. Selain itu, ada jaringan tambahan yang mendeteksi artefak spesifik dari teknik manipulasi tertentu, sehingga membuat arsitektur ini lebih andal dibandingkan metode berbasis artefak lainnya.

4.7. A Convolutional LSTM based Residual Network for Deepfake Video Detection

Penelitian ini [18] menggunakan *Convolutional LSTM* (ConvLSTM) sebagai bagian utama dari arsitektur CLRNNet. ConvLSTM digunakan untuk menangani data spasial-temporal, di mana input berupa urutan frame video yang berurutan. Selain itu, transfer learning juga diterapkan untuk meningkatkan generalisasi pada metode *Deepfake* yang tidak ada dalam data pelatihan.

Pada dataset DeepFake (DF), akurasi CLRNNet mencapai 99,02%, namun setelah melakukan transfer learning, akurasi turun menjadi 90,95%. Pada dataset FaceSwap (FS) setelah transfer learning, akurasi meningkat dari 50,00% menjadi 83,08%. Tidak ada informasi spesifik mengenai waktu pelatihan atau inferensi yang diberikan dalam penelitian ini. Jumlah Epoch yang digunakan 150

Model ini menerapkan *transfer learning* sebagai strategi optimasi untuk meningkatkan performa deteksi pada metode *Deepfake* yang tidak ada dalam data pelatihan. Ini membantu model mendeteksi

Deepfake dengan hanya memerlukan sedikit sampel dari metode baru. Arsitektur CLRNNet menggunakan ConvLSTM untuk mengidentifikasi artefak antar frame dalam urutan video, yang merupakan tambahan dibandingkan metode konvensional. Selain itu, *transfer learning* diterapkan untuk meningkatkan kemampuan generalisasi pada metode *Deepfake* baru

4.8. Spatio-temporal Features for Generalized Detection of Deepfake Videos

Penelitian ini [19] menggunakan CNN 3D untuk memodelkan hubungan spasial-temporal dalam video. CNN 3D memungkinkan pemrosesan fitur spasial dan temporal secara bersamaan, berbeda dengan model deteksi berbasis gambar (CNN 2D) atau model sekuensial (RNN). Hasil akurasi eksperimen menunjukkan performa pada dataset Deeper Forensics, dengan model I3D memberikan hasil terbaik (Xcept 24.28, E-B3 16.43, RNN 18.57, R3D 56.43, I3D 78.57), di mana I3D menunjukkan kinerja tertinggi. CNN 3D adalah arsitektur utama yang digunakan untuk menangkap fitur spasial-temporal, berbeda dari pendekatan tradisional yang menggunakan CNN 2D atau RNN.

4.9. WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection

Menurut penelitian [20] Metode ini menggunakan kombinasi *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) untuk menganalisis fitur spasial dan temporal dalam bingkai video. Komponen CNN mengekstrak fitur dari setiap frame, sedangkan komponen RNN menangkap dinamika temporal di seluruh frame dan mendapatkan akurasi 85.50%.

4.10. Deepfake Video Detection through Optical Flow based CNN

Penelitian ini [21] menggabungkan beberapa metode, yaitu: Optical Flow untuk mengidentifikasi gerakan antar frame video dan menemukan perbedaan antara video asli dan video yang dimanipulasi. CNN (Convolutional Neural Network), menggunakan arsitektur pra-latih seperti VGG16 dan ResNet50 sebagai backbone untuk memproses fitur optical flow. Transfer Learning untuk mengatasi masalah dataset yang relatif kecil dengan menerapkan transfer learning pada beberapa lapisan awal jaringan.

Hasil akurasi yang diperoleh untuk dua arsitektur CNN yang diuji adalah: VGG16: 81,61% ResNet50: 75,46%. Penggunaan *transfer learning* sebagai metode optimasi untuk memperbaiki performa model mengingat ukuran dataset yang kecil. Ini memungkinkan model untuk memanfaatkan pengetahuan dari jaringan pra-latih.:

Penelitian ini menggunakan CNN dengan arsitektur pra-latih (VGG16 dan ResNet50), yang merupakan tambahan dibandingkan dengan model CNN yang dibangun dari awal. Penggunaan Optical Flow juga merupakan pendekatan tambahan yang

membantu model mendeteksi perbedaan gerakan antar frame video, meningkatkan efektivitas deteksi manipulasi.

4.11. Analisis Gambar Wajah Palsu: Mendeteksi Keaslian Gambar Yang Dimanipulasi Menggunakan Metode Variational Autoencoder Dan Forensics Deep Neural Network

Penelitian ini [22] menggunakan Variational Autoencoder (VAE) untuk mendeteksi keaslian gambar dengan mempelajari representasi dari gambar asli dan gambar yang telah dimanipulasi. Akurasi Prediksi pertama menunjukkan kepercayaan 50,46% untuk gambar yang terdeteksi sebagai 'fake'. Prediksi kedua menunjukkan kepercayaan 93,17% untuk gambar yang terdeteksi sebagai 'fake'. Tidak ada informasi spesifik mengenai waktu pelatihan atau inferensi yang diberikan dalam penelitian ini. Jumlah epoch yang digunakan 30 Penelitian tidak secara eksplisit menyebutkan optimasi, tetapi penggunaan VAE untuk menghasilkan model probabilistik sebagai pendekatan optimasi untuk membedakan antara gambar asli dan gambar yang dimanipulasi. Arsitektur yang digunakan adalah VAE, yang merupakan pendekatan probabilistik dalam generasi data dan pengenalan pola.

4.12. Dynamic texture analysis for detecting fake faces in video sequences

Penelitian ini [20] menggunakan Support Vector Machine (SVM) untuk mengklasifikasikan berbagai teknik manipulasi wajah, termasuk: DeepFake (DF) Face2Face (F2F) FaceSwap (FSW) FaceShifter (FSH) Neural Textures (NT) Akurasi dalam skenario manipulasi tunggal: Algoritma untuk DF dan FSW mencapai akurasi lebih dari 90%. Algoritma untuk F2F dan NT memiliki akurasi di bawah 85%. Dalam skenario manipulasi ganda, akurasi rata-rata turun menjadi sekitar 70%, dengan peningkatan tingkat positif palsu yang lebih tinggi. Penelitian ini tidak menyebutkan secara eksplisit tentang optimasi. Penelitian ini menggunakan SVM, yang merupakan metode klasifikasi yang umum, tanpa menyebutkan adanya tambahan arsitektur yang signifikan

4.13. Deepfake-Image Anti-Forensics with Adversarial Examples Attacks

Penelitian ini [21] menggunakan algoritma yang disebut Poisson noise DeepFool (PNDF), yang dirancang untuk menyerang detektor deepfake dengan menambahkan gangguan pada gambar. Akurasi sebelum gangguan ditambahkan, detektor memiliki Area Under Curve (AUC) sebesar 0.9999 dan akurasi deteksi gambar palsu (f-ac) sebesar 0.9997. Setelah menerapkan PNDF, AUC turun menjadi 0.0331 dan f-ac turun menjadi 0.0731. Ini menunjukkan penurunan drastis dalam kemampuan detektor untuk mengenali gambar palsu.

4.14. Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network

Penelitian ini [22] menggunakan Convolutional Neural Network (CNN) sebagai metode utama untuk mendeteksi dan mengklasifikasikan citra. Hasil akurasi model untuk data training adalah 99,56% (0,9956). Hasil akurasi untuk data validasi adalah 87,06% (0,8706). Penelitian ini tidak secara eksplisit menyebutkan teknik optimasi tertentu. Namun, arsitektur CNN yang digunakan secara inheren merupakan bentuk optimasi dalam proses ekstraksi fitur dan klasifikasi citra secara bersamaan. Arsitektur CNN yang digunakan terdiri dari tiga lapisan utama: Konvolusi layer untuk mengekstrak fitur-fitur penting dari data. Pooling layer untuk mengurangi dimensi peta fitur, yang membantu mengurangi kompleksitas dan meningkatkan efisiensi

4.15. Trans-DF: A Transfer Learning-based end-to-end Deepfake Detector

Penelitian ini [8] menggunakan Generative Adversarial Networks (GANs) untuk mendeteksi dan mengklasifikasikan gambar dan video yang dimanipulasi (deepfake). Akurasi Model yang diusulkan dalam makalah ini mencapai akurasi 90,2% dalam mengklasifikasikan gambar palsu dari yang asli

5. KESIMPULAN

Dari berbagai literatur metode deteksi deepfake menunjukkan bahwa Deep Convolutional Generative Adversarial Network (DCGAN) merupakan metode unggulan dengan akurasi mencapai 100% pada iterasi ke-10 dari 8000 epoch. DCGAN tergolong ke dalam computer vision karena sering digunakan untuk menghasilkan dan menganalisis gambar. DCGAN menggabungkan konsep dari Generative Adversarial Networks (GANs), yang terdiri dari dua jaringan (generator dan discriminator) yang saling bersaing untuk menghasilkan data baru yang mirip dengan data asli. Dengan demikian, computer vision adalah bidang yang paling cocok untuk kasus deepfake, dan penerapan teknik optimizer seperti Adam dan Binary Cross-Entropy (BCE) dapat lebih meningkatkan kinerja deteksi, menjadikan metode ini lebih efektif dalam menghadapi tantangan manipulasi gambar maupun video.

Mengingat keberhasilan DCGAN dalam mendeteksi deepfake dengan akurasi yang tinggi, penerapan metode ini sebaiknya diperluas, terutama pada aplikasi yang berkaitan dengan analisis gambar dan video. Penggunaannya dapat diperkuat di berbagai sektor seperti keamanan digital, forensik, dan media sosial untuk memerangi penyebaran konten deepfake.

DAFTAR PUSTAKA

- [1] M. Albahar and J. Almalki, "DEEPFAKES : THREATS AND COUNTERMEASURES SYSTEMATIC REVIEW," vol. 97, no. 22, pp. 3242–3250, 2019.
- [2] D. K. Citron, R. Chesney, B. Chesney, and D. Citron, "Deep Fakes : A Looming Challenge for Privacy , Democracy , and National Security," 2019.
- [3] R. Andreas, D. Cozzolino, L. Verdoliva, J. Thies, M. Nießner, and C. Riess, "FaceForensics ++ : Learning to Detect Manipulated Facial Images".
- [4] P. Yu, "A Survey on Deepfake Video Detection," no. October 2020, pp. 607–624, 2021, doi: 10.1049/bme2.12031.
- [5] D. Wodajo and S. Atnafu, "Deepfake Video Detection Using Convolutional Vision Transformer," 2021, [Online]. Available: <http://arxiv.org/abs/2102.11126>
- [6] F. Mahmud, Y. Abdullah, M. Islam, and T. Aziz, "Unmasking Deepfake Faces from Videos Using An Explainable Cost-Sensitive Deep Learning Approach," 2023 26th Int. Conf. Comput. Inf. Technol. ICCIT 2023, no. December, pp. 13–15, 2023, doi: 10.1109/ICCIT60459.2023.10441026.
- [7] V. L. L. Thing, "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers," *Proc. 2023 IEEE Int. Conf. Cyber Secur. Resilience, CSR 2023*, pp. 246–253, 2023, doi: 10.1109/CSR57506.2023.10225004.
- [8] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision : A Brief Review," vol. 2018, 2018, doi: 10.1155/2018/7068349.
- [9] I. Goodfellow, "Deep Learning".
- [10] M. Westerlund, "The Emergence of Deepfake Technology : A Review," 2019.
- [11] R. A. S. Rahayu and Handri Santoso, "ANALISIS GAMBAR WAJAH PALSU : MENDETEKSI KEASLIAN GAMBAR YANG DIMANIPULASI MENGGUNAKAN METODE VARIATIONAL AUTOENCODER DAN FORENSICS DEEP NEURAL NETWORK ANALYSIS OF FAKE FACE IMAGES : DETECTING THE AUTHENTICITY OF MANIPULATED IMAGES USING VARIATIONAL AUTOE," vol. 2, no. 9, pp. 2701–2726, 2023.
- [12] M. Kumar and H. K. Sharma, "A GAN-Based TabModel of Deepfake Detection in Data Social Media Kumar in Model of Deepfake Detection Social Media," *Procedia Comput. Sci.*, vol. 218, pp. 2153–2162, 2023, doi: 10.1016/j.procs.2023.01.191.
- [13] M. Indra, I. Nurtanio, and A. Achmad, "Deepfake detection in videos using Long Short-Term Memory and CNN ResNext," vol. 14, no. 3, pp. 178–185, 2022.
- [14] P. Kawa and P. Syga, "A Note on Deepfake Detection with Low-Resources".
- [15] L. Guarnera, O. Giudice, and S. Battiato, "Fighting Deepfake by Exposing the Convolutional Traces on Images".
- [16] X. Li, Y. Chen, S. Wang, Q. Lu, and A. Group,

- “Sharp Multiple Instance Learning for DeepFake Video Detection,” 2020.
- [17] Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, “DeepFake Detection Based on Discrepancies Between Faces and their Context,” pp. 1–10.
- [18] S. Tariq, S. Lee, and S. S. Woo, “A Convolutional LSTM based Residual Network for Deepfake Video Detection”.
- [19] I. Ganiyusufoglu and L. M. Ng[^], “Spatio-temporal Features for Generalized Detection of Deepfake Videos,” pp. 1–11, 2020.
- [20] B. Zi, M. Chang, J. Chen, and Y. Jiang, “WildDeepfake : A Challenging Real-World Dataset for Deepfake Detection,” 2020.
- [21] I. Amerini, L. Galteri, R. Caldelli, A. Del Bimbo, and M. Integration, “Deepfake Video Detection through Optical Flow based CNN,” no. Micc, pp. 2019–2021, 2019, doi: 10.1109/ICCVW.2019.00152.
- [22] J. V. C. I. R, M. Bonomi, C. Pasquini, and G. Boato, “Dynamic texture analysis for detecting fake faces in video sequences ☆,” *J. Vis. Commun. Image Represent.*, vol. 79, no. March 2020, p. 103239, 2021, doi: 10.1016/j.jvcir.2021.103239.